



Computational Reconstruction of Macromolecular Assemblies

A. Griewel, M. Rarey

published in

From Computational Biophysics to Systems Biology (CBSB07),
Proceedings of the NIC Workshop 2007,
Ulrich H. E. Hansmann, Jan Meinke, Sandipan Mohanty,
Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. 36, ISBN 978-3-9810843-2-0, pp. 121-123, 2007.

© 2007 by John von Neumann Institute for Computing

Permission to make digital or hard copies of portions of this work for personal or classroom use is granted provided that the copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise requires prior specific permission by the publisher mentioned above.

<http://www.fz-juelich.de/nic-series/volume36>

Computational Reconstruction of Macromolecular Assemblies

Axel Griewel and Matthias Rarey

Center for Bioinformatics
University of Hamburg
Bundesstr. 43
20146 Hamburg, Germany
E-mail: {griewel, rarey}@zbh.uni-hamburg.de

Cryo electron microscopy (cryo EM) has proven to be an effective technique to gather low resolution information of macromolecular structures. Methods like X-ray crystallography and NMR spectroscopy on the other hand supply atomic detail structures but tend to fail in the investigation of large specimen. We present a method for elucidating macromolecular structures by combining atomic detail structures of subunits with overall shape information gathered by cryo EM – a problem for which only few computational methods have been published so far. In our approach atomic detail structures are reduced to a set of spheres by computational methods like gift-wrapping and Delaunay triangulation. These describing spheres are then matched to a hard sphere-filtered cryo EM map employing pose clustering, a technique adapted from computer vision. Tests were performed using experimental and simulated cryo EM maps. Calculations yield results within 3.5 Å rmsd accuracy or better in a runtime of less than 5 minutes.

1 Introduction

Cryo electron microscopy is as a powerful imaging technique for the elucidation of macromolecular structures. The resolution of the resulting density maps, however, does not allow for atomic detail interpretation. For this task methods like X-ray crystallography and NMR spectroscopy are more suited, while they also tend to fail in the investigation of large specimen. Combining atomic detail structures of subunits with cryo EM is a viable method for determining macromolecular structures. We developed a fast and accurate tool incorporating information from different sources for the computational reconstruction of macromolecular assemblies (CRoMA).

All published methods for this problem except one rely on a sequential scan of the six dimensional degrees of freedom given by rotation and translation of the two objects relative to each other. Only one published approach utilizes topology representing neural networks. Three approaches for a scoring function can be found in literature. One is to calculate a synthetic cryo EM map of the atomic structure at the desired resolution and to determine the cross correlation of the two maps. Another approach utilizes laplacian filtered maps yielding a score reflecting the overlap of borders. A third scoring function uses trilinear interpolation for each atom.^{1,2}

2 Method

The atomic detail structure as well as the map are preprocessed yielding descriptors consisting of spheres of different radii. These descriptors are utilized to determine favourable placements by a technique adapted from computer vision.

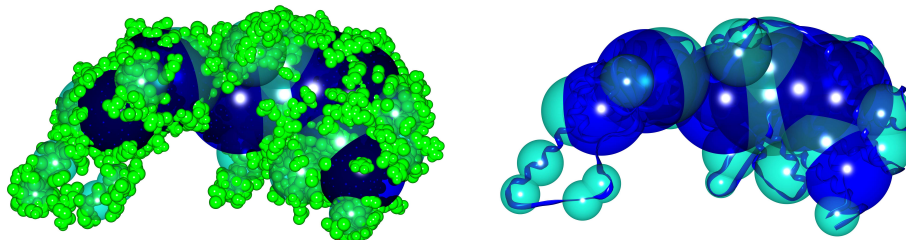


Figure 1. Atomic detail structure and ribbon model of a monomer of the thermosome from *T. Acidophilum* (PDB accession code 1A6D) together with the implemented sphere-descriptor

The goal of the descriptor for structures given with atomic resolution is to determine few large spheres describing the shape of the structure in decent detail. Therefore all surface atoms of the structure are determined using a gift-wrapping algorithm. For the set of surface atoms the three-dimensional Delaunay triangulation is computed. An essential property of this triangulation is that each tetrahedron's circumsphere does not contain any other surface atom, i. e. also the largest spheres inside the structure are contained in this set of spheres. All circumspheres outside the structure are discarded while a subselection of spheres describing the interior of the structure is chosen employing a set cover algorithm. An example of a descriptor can be found in Figure 1. The cryo EM map descriptor is constructed by filtering with isodense spheres of different radii. The radii for filtering are chosen according to the size of the spheres found as structure descriptor.

For calculating the placement of the high resolution structure in the low resolution density map, pose clustering³ – a powerful method from pattern recognition – is used. Triangles between centers of characteristic spheres in the structure-descriptor are used to query a database of triangles build up by selected voxels of the filtered map. The pose clustering algorithm then proposes positions in which the three structure spheres generating the triangle coincide with map spheres of similar radius.

All proposed placements are subject to a quick scoring function which utilizes the sphere representation of structure and map. A certain amount of proposed placements is then retained and scored according to a fine-grained scoring function based on trilinear interpolation. The resulting placements can be clustered given a user defined rmsd threshold.

3 Results

A first test was performed by docking monomers taken from the molecular complex GroEL deposited in the PDB with accession code 1OEL to map 1081 taken from the EM-Search database⁴. The rmsd between the calculated and the correct position was 1.7 Å and the computing time was less than four minutes.

Further tests were conducted on simulated maps using various structures. The result of docking structures to urate oxidase (2FUB) is shown in Figure 2. Within ten minutes run-time placements with 3.4 Å rmsd to the original placement were found. Another example

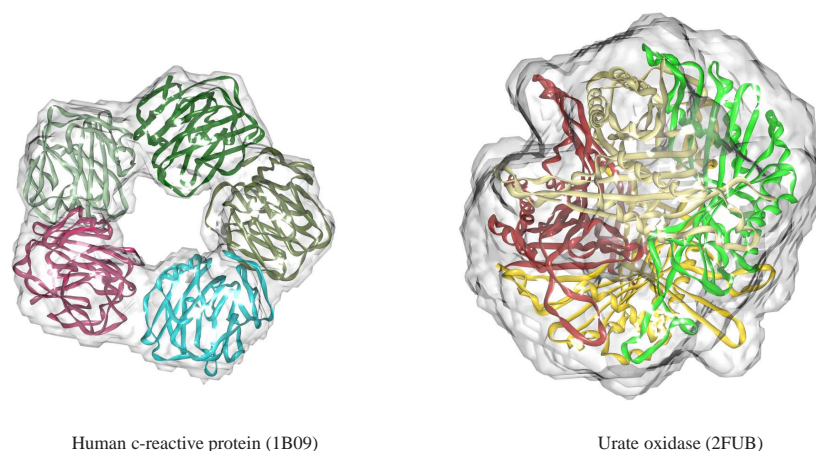


Figure 2. Solution structures

is docking monomers to a map generated from human c-reactive protein (1B09). Here a rmsd of 2.3 Å was achieved in a runtime below one minute. The resulting placements for the latter two examples can be found in Figure 2.

4 Perspectives

Besides the success in several cases, the algorithm fails in a few docking scenarios. This is the case, if the molecular mass of the structure to be docked makes up only a very small fraction of the overall mass. Further improvements to the scoring function will help tackling these cases. Although not yet fully developed, first applications of this approach suggest that the algorithm is going to be highly efficient and effective.

References

1. J. Frank, *Three-Dimensional Electron Microscopy of Macromolecular Assemblies* (Oxford University Press, New York, 2006).
2. M. G. Rossmann, Marc C. Morais, Petr G. Leiman, and Wei Zhang, *Combining X-ray Crystallography and Electron Microscopy*, *Structure* **13**, 355–362, 2005.
3. S. Linnainmaa, D. Harwood, L. S. Davis, *Pose determination of a three-dimensional object using triangle pairs*, *IEEE trans. pattern anal. mach. intell.* **10**, 634–647, 1988.
4. *A database for EM*, *Nat Struct Biol* **10**, 313, 2003.